



**European Language
Resource Coordination**
Connecting Europe Facility

ELRC Workshop Report for Austria



Author(s): Gerhard Budin, University of Vienna
Dissemination Level: Public
Version No.: <V1.1>
Date: 2016-05-30



Contents

1	<u>Executive Summary</u>	3
2	<u>Workshop Agenda</u>	4
3	<u>Summary of Content of Sessions</u>	5
3.1	Session 1: Opening and Welcome	5
3.2	Session 2: Welcome on behalf of the European Commission and the Austrian government	5
3.3	Session 3: Aims and Objectives of CEF.AT Initiative	5
3.4	Session 4: Europe and Multilingualism	5
3.5	Session 5: How can Public Institutions benefit from the CEF.AT Platform?	5
3.6	Session 6: Languages and Language technologies in Austria	5
3.7	Session 7: “Automated Translation”: How does it work?	6
3.8	Session 8: What Data are needed in order to optimize CEF.AT?	6
3.9	Session 9: Legal Framework for Contributing Data for the optimization process	6
3.10	Session 9: Data and Language Resources: Technical Aspects and Best Practices	6
3.11	Session 9: Summary and Conclusions, Next steps and Closing remarks	6
4	<u>Synthesis of Workshop Discussions</u>	7
4.1	Panel 1: Multilingual Public Services in Austria	7
4.2	Panel 2: Data and language resources in Austria	7
4.3	Discussion: How can the Public Sector contribute, how can we engage and how can we benefit from CEF.AT?	7
5	<u>Workshop Presentation Materials</u>	8

1 Executive Summary

The Austrian ELRC workshop took place on the 15th of April 2016 at the Centre for Translation Studies at the University of Vienna. About 36 participants attended the workshop. Simultaneous interpretation between English and German in both directions was provided all day. In 11 sessions and 3 panels/discussions the most important data on the CEF.AT process and how Austrian public services can participate were presented and discussed. This included the policy of European Multilingualism and the role of machine translation in the endeavor to provide multilingual services to all citizens in the European Union. The language situation in Austria and the state-of-the-art of language technologies industry in Austria were explained, as well as the principles of statistical machine translation and the role of language resources (parallel corpora, terminologies, etc.) in providing a better quality of the use of such translation technologies. The 3 panels and discussions provided useful information on the situation of the public translation services in Austria and what the issues are concerning the use of automated translation. Practical issues and the next steps for the identification and provision of relevant data from the Austrian public services for the ELRC process were discussed and explained. The readiness to participate and the positive expectation of a benefit of using CEF.AT were clearly expressed. Details about the event, including workshop presentations, can be found at the dedicated webpage, at: <http://lr-coordination.eu/de/austria>.

2 Workshop Agenda

Agenda for Austria ELRC Training Workshop, 15th of April 2016

Venue:

Centre for Translation Studies, University of Vienna, 1st floor, Hörsaal 1

Gymnasiumstraße 50, 1190 Vienna

08:00 – 09:00 Registration, Coffee

09:00 – 09:05 Opening and Welcome (ELRC team: Stelios Piperidis, ILSP & Gerhard Budin, Univ. Vienna, Technology NAP)

09:05 – 09:20 Welcome on behalf of the European Commission and the Austrian government (Achim Braun, DGT Office in Austria & Bruno Nestler, BMLVS, Public Sector NAP)

09:20 – 09:35 Aims and Objectives of ELRC (Stelios Piperidis & Gerhard Budin)

09:35 – 09:50 Europe and Multilingualism (Achim Braun)

09:50 – 10:20 How can Public Institutions benefit from the CEF.AT Platform? (Daniel Klivanec, European Commission)

10:20 – 10:50 Languages & Language technology in Austria (Vesna Lusicky, Univ. Vienna)

10:50 – 11:20 Coffee break and Networking

11:20 – 12:00 Panel: Multilingual Public Services in Austria (Elisabeth Taucher, BMI, Sonja Schnitzer, BMLFUW, Ingrid Haussteiner, OeNB, Moderator: Achim Braun)

12:00 – 12:30 “Automated Translation”: How does it work? (Bartholomäus Wloka, Univ. Vienna)

12:30 – 13:30 Lunch Break

13:30 – 14:00 What Data is needed? (Bartholomäus Wloka, Univ. Vienna)

14:00 – 14:30 Legal Framework for Contributing Data (Christian Galinski, Infoterm)

14:30 – 15:00 Panel: Data and language resources in Austria (Elisabeth Taucher, BMI, Sonja Schnitzer, BMLFUW, Ingrid Haussteiner, OeNB, Moderator: Jürgen Kotzian, LVAK/SIB, ARG GUT)

15:00 – 15:30 Coffee break and Networking

15:30 – 16:00 Data and Language Resources: Technical Aspects and Best Practices (Stelios Piperidis)

16:00 – 16:30 Discussion: How can the Public Sector contribute, how can we engage and how can we benefit from CEF.AT? (Gerhard Budin, Achim Braun, Stelios Piperidis)

16:30 – 16:45 Summary and Conclusions, Next steps and Closing remarks (Stelios Piperidis, Bruno Nestler, Gerhard Budin)

3 Summary of Content of Sessions

3.1 Session 1: Opening and Welcome

ELRC team representative Stelios Piperidis (ILSP) and National Technology Anchor Point Gerhard Budin (Univ. Vienna) opened the workshop and welcomed the participants.

3.2 Session 2: Welcome on behalf of the European Commission and the Austrian government

Achim Braun, EU Representation in Austria, DG Translation and Bruno Nestler, BMLVS, Public Sector NAP, welcomed the workshop participants on behalf of the EU Commission and on the behalf of the Austrian government.

3.3 Session 3: Aims and Objectives of CEF.AT Initiative

Stelios Piperidis explained the workshop aims and the objectives in detail. The workshop objectives included raising awareness on the value and importance of data held by public services in overcoming language barriers, engaging us and our institutions in the identification and sharing of data for CEF.AT, helping with legal and technical issues associated with the collection and/or provision of data by public service administrations, and sharing data for the best results for your needs and your language with European Commission. Modern MT is all about data, it learns how to translate from human translations (i.e. data). The best data to improve MT and for CEF.AT are the data of ALL of us.

In each country ELRC is working with national experts (National Anchor Points) and DGT representatives, namely in Austria:

Technology NAP: Gerhard Budin (Vienna University)

Public Sector NAP: Bruno Nestler (National Defense Academy, Austrian Armed Forces)

Local DGT offices: Achim Braun, DG Translation, Vienna Office

The common objective is to share data for the best results for all our needs and our language (German).

3.4 Session 4: Europe and Multilingualism

Achim Braun described the language policy of the European Union and pointed out the importance of multilingualism for every citizen in every area of society.

3.5 Session 5: How can Public Institutions benefit from the CEF.AT Platform?

Daniel Klivanec (European Commission) explained how the EC Machine Translation Service works, which role it plays in CEF as a whole and that in turn in the context of the DSM initiative for a digital single market. He also explained how the national translation services can participate and how they will benefit from the CEF.AT platform.

3.6 Session 6: Languages and Language technologies in Austria

Vesna Lusicky (Univ. Vienna) discussed the official language in Austria (German national variety for Austria) as well as regional official languages (at the same time minority languages (Burgenland Croatian, Romani, Slovakian, Slovenian, Czech, and Hungarian) and the

ELRC Workshop Report for <country>

Austrian Sign Language. She also stressed the rising need for translation in Austria due to increases in regional development and cooperation, traffic and mobility, energy and climate and environment protection, trans-national legal and administrative affairs, tourism, trade and commerce, research cooperation, etc. Vesna Lusicky also provided statistical numbers to underpin the importance of multilingualism in Austria. She then elaborated on the important role of the language technology landscape in Austria, both in academic institutions as well as in the industry, and listed the research projects funded by the EC in the area of language technologies.

3.7 Session 7: “Automated Translation”: How does it work?

Bartholomäus Wloka (Univ. Vienna) explained the basic principles of machine translation, in particular of the statistical MT paradigm and its further development into phrase-based SMT.

3.8 Session 8: What Data are needed in order to optimize CEF.AT?

Bartholomäus Wloka (Univ. Vienna) explained in detail which data are needed for optimizing CEF.AT and where these data can be found. He also explained which data formats and which types of data are preferable and how the data pipeline works.

3.9 Session 9: Legal Framework for Contributing Data for the optimization process

Christian Galinski (Infoterm) focused in his presentation on the legal framework under which data can be provided for the CEF.AT optimization process. Some basic distinctions have to be made, between different kinds of content, different legal barriers and which kind of contractual agreement as well as licensing scheme is most appropriate for which kind of data. In addition to a checklist of what needs to be taken into account Christian Galinski also dealt with recent and ongoing legislation and their impact as well as future trends concerning IPR in the modern information society.

3.10 Session 9: Data and Language Resources: Technical Aspects and Best Practices

Stelios Piperidis explained in detail the technical details of how data are taken over for the integration work (data pipeline) and how the ELRC team can help in this process of data preparation, data transfer and description of licensing conditions, how to upload the data on the ELRC repository, etc.

3.11 Session 9: Summary and Conclusions, Next steps and Closing remarks

Stelios Piperidis, Bruno Nestler and Gerhard Budin summarized the contents and major outcomes of the workshop and repeated the next steps to be taken. They thanked the participants for the active participation.

4 Synthesis of Workshop Discussions

4.1 Panel 1: Multilingual Public Services in Austria

Achim Braun served as moderator of this panel; the panel members were Elisabeth Taucher, BMI, Sonja Schnitzer, BMLFUW, Ingrid Haussteiner, OeNB. The issues the panel discussed included multilingualism in public services in Austria, about the current situation, which languages are most needed and dealt with, whether there are any major trends, whether machine translation is already used, which problems are encountered when using it, which other needs have been identified, and which kinds of documents are typically translated.

It became clear that the situation is quite different from one institution to another. The Ministry of Interior Affairs has primarily highly sensitive texts to be translated; in the Austrian National Bank on the other hand there is already a close cooperation with the European Central Bank (which is already using MT@EC). None of the panelists have already worked with MT@EC but they are interested in starting to use it and they consider the ELRC initiative to be very useful, also for themselves in the near future.

4.2 Panel 2: Data and language resources in Austria

The moderator of the second panel was Jürgen Kotzian from the Austrian Defense Academy/Ministry of Defense and the panelists were identical to the ones in panel 1: Elisabeth Taucher, BMI, Sonja Schnitzer, BMLFUW, Ingrid Haussteiner, OeNB.

The key questions were: which language pairs are relevant, advantages and disadvantages of machine translation vs. translation memory, does the inhouse translation service cover all the internal translation needs, is there quality management, are the language resources of the public service usable for CEF.AT? The key questions were where the relevant resources are and in which quality? What are the pre-requisites for data sharing with the EU Commission? Which steps are necessary?

The language pair by far most often used is German-English/English-German. There are relevant language resources of different types in the language services of the panelists that can be shared; the CEF.AT service will definitely be useful for them. The question is in which format the data have to be provided. All these details will then be dealt with in detail at the individual level of negotiation.

4.3 Discussion: How can the Public Sector contribute, how can we engage and how can we benefit from CEF.AT?

Gerhard Budin, Achim Braun, and Stelios Piperidis moderated this discussion session. Building upon the two panels it was confirmed that a number of institutions are willing and ready to share resources for the ELRC initiative and that there is quite some interest in using CEF.AT in the future. ELRC will get in contact with all the relevant representatives of the public services for the next steps. The Austrian coordination group of public language services (ARGE GUT) has already started to serve as a very useful information pool for this purpose.

5 Workshop Presentation Materials

All workshop presentations are uploaded to the dedicated workshop webpage, at http://lr-coordination.eu/de/austria_agenda.